

MASCHINELLE ÜBERSETZUNGEN AUS DEM JAPANISCHEN

LEXIKALISCHE, SYNTAKTISCHE UND SEMANTISCHE AMBIGUITÄTEN

Barbara MANTHEY

Die maschinelle Übersetzung (in Fachkreisen übliche Abkürzung: MÜ) ist ein multidisziplinärer Forschungszweig der Computerlinguistik, an dem – nach einer deutlichen ingenieurwissenschaftlichen Dominanz bis Ende der siebziger Jahre – heutzutage nicht nur Computerlinguisten, Informatiker und Ingenieurwissenschaftler beteiligt sind, sondern auch Linguisten, Kognitionsforscher, Übersetzer und Übersetzungswissenschaftler sowie Psychologen. Im Rahmen von Veranstaltungen im Bereich der (literarischen) Humanübersetzung hat die Problematik der maschinellen Übersetzung seit jeher einen schweren Stand. Hauptgrund war (und ist noch immer) die mangelnde Kommunikation und Kooperation zwischen den beteiligten Disziplinen und das damit einhergehende fehlende Verständnis für die jeweils unterschiedlichen Standpunkte und Herangehensweisen. Während Kritik an der Qualität von maschinellen Übersetzungssystemen jedoch in Europa – und hier speziell in Deutschland – deutlich formuliert und in der Forschung aufgegriffen wird, findet Forschung und Entwicklung in Japan nach wie vor in einem computerlinguistisch und ingenieurwissenschaftlich dominierten Vakuum ohne nennenswerten interdisziplinären Austausch statt.

Das Japanese Technology Evaluation Center (JTEC) veröffentlichte 1992 eine komparative Analyse des MÜ-Bereichs in Japan und den USA und kam zu dem Schluß, daß die finanzielle Förderung der Forschung und Entwicklung im Bereich maschineller Übersetzung in Japan bereits Anfang der neunziger Jahre substantiell höher war als in den USA. Auch Bestrebungen zur Kommerzialisierung von maschineller Übersetzungssoftware gibt es in den USA nicht in diesem Umfang, da dort die Entwicklungszeiten für neue maschinelle Übersetzungssysteme – ebenso wie in Europa – verhältnismäßig lang sind und die Ergebnisse aus Forschung und Entwicklung aus diesem Grund erst seit kurzem mit einer Kommerzialisierung in Verbindung gebracht werden (CARBONELL 1992: 12–13). Entsprechend der politischen, ökonomischen und gesellschaftlichen Erfordernisse Japans konzentrierten sich die japanischen Aktivitäten im Bereich der maschinellen Übersetzung größtenteils auf Englisch und Japanisch.

Anfang der neunziger Jahre existierten bereits über zwanzig Systeme für Übersetzungen zwischen diesen beiden Sprachen. Der Schwerpunkt der Bemühungen lag zunächst auf Übersetzungen aus dem Japanischen ins Englische wegen der weit größeren Nachfrage für diese Übersetzungsrichtung. Zugleich nahm die Motivation, Systeme für Übersetzungen vom Englischen ins Japanische zu entwickeln, stark zu. Während im Fall der Übersetzungen ins Englische japanische Informationen einem breiten ausländischen Markt zugänglich gemacht werden können, sprechen verschiedene Gründe für eine verstärkte Entwicklung von maschinellen Übersetzungssystemen für die Übersetzung ins Japanische: Zum einen besteht in Japan eine große Nachfrage nach Übersetzungen englischsprachiger Informationen, zum anderen ist die maschinelle Analyse der Ausgangssprache Englisch im Vergleich zum Japanischen wesentlich einfacher. Hinzu kommt, daß für die Zielsprache Japanisch eine erheblich größere Anzahl von Posteditoren zur Verfügung steht, deren Schulung und Anstellung wesentlich geringere Kosten verursacht als die Schulung und Anstellung von englischen Muttersprachlern. Seit Beginn der neunziger Jahre ist man in der Forschung und Entwicklung jedoch dazu übergegangen, auch andere Sprachen einzubeziehen. Das Center for the International Cooperation in Computerization (CICC), ein vom Ministry of International Trade and Industry (MITI) organisiertes internationales Konsortium unter Beteiligung von sieben bedeutenden Industrieunternehmen, entwickelt bereits seit 1987 ein interlinguabasiertes maschinelles Übersetzungssystem für die Sprachen Japanisch, Chinesisch, Thailändisch, Malaiisch und Indonesisch (VASCONCELLOS 1992: 41–46).¹

Bei der Kommerzialisierung bereits entwickelter maschineller Übersetzungssysteme fällt Japan heute eine internationale Schlüsselposition zu. 22 japanische Unternehmen vertreiben inzwischen rund einhundert-zehn Produkte im Bereich der maschinellen Übersetzung allein für die Übersetzung in die bzw. aus der japanischen Sprache (AAMT 1998: 18–23). Die Kommerzialisierung von maschineller Übersetzungssoftware „für den Hausgebrauch“ begann in Japan 1990 aufgrund einer technischen Verbesserung im PC-Bereich (32-Bit-Shift): Die bislang nur auf Mainframe-Geräten oder Workstations installierbaren maschinellen Übersetzungssysteme konnten nun auch als PC-Versionen vertrieben werden. Kostete die erste Software („PC-Transer/ej“ der Firma Nova) noch 248 000 Yen, kollabierten die Preise bereits im November 1994, als Catena die Software „Korya Eiwa“ für 9 800 Yen auf den Markt brachte. Noch 1994 ging man in Japan davon aus, pro Monat lediglich einige

¹ CICC-Projektberichte finden sich z.B. bei KAWASHIMA (1997: 18) und KAMIGANE (1997: 29–33).

tausend dieser Softwarepakete verkaufen zu können, doch infolge des immer stärkeren Übersetzungsbedarfs der wachsenden Internet-Gemeinde wurde diese Schätzung ab 1996 um ein Vielfaches übertroffen (KOGA 1996: 6).²

Den Produkten der einzelnen Unternehmen, zu denen die großen Elektronik- und Computerhersteller wie Hitachi, NEC, Fujitsu, Oki, Sharp, Toshiba, Matsushita und IBM Japan ebenso gehören wie auf maschinelle Übersetzungssoftware spezialisierte Unternehmen (z.B. CSK, Catena, Nova), liegt zwar in der Regel die gleiche, in den achtziger Jahren entwickelte Technologie zugrunde (TANAKA 1996: 2–3), aber sie unterscheiden sich in folgenden Bereichen: erforderliches Betriebssystem, bereitgestellte Sprachenpaare, Übersetzungsrichtung, Umfang des mitgelieferten Basiswörterbuchs, notwendiger Speicherplatz, Verarbeitungsgeschwindigkeit und Preis.

Betriebssystem:\nSprachen:	Windows	Mac OS	MS-Dos	Unix	Sonstige
E-J	54	9	1	4	2
J-E	29	6		4	
F-J	2				
D-J	1				
J-K	1		1		
K-J	1		1		
Ch-J	1				
J-Ch			1		
I-J	1				

Tab.: Maschinelle Übersetzungssysteme auf dem japanischen Markt (nach Sprachenpaaren)

E=Englisch, J=Japanisch, F=Französisch, D=Deutsch, K=Koreanisch, Ch=Chinesisch, I=Italienisch;

Mehrfachnennungen bei Hybrid-Systemen möglich.

Quelle: Eigene Berechnungen nach AAMT (1998: 18–23).

Das in den vergangenen zwei Jahrzehnten stark zunehmende Interesse an elektronischen Hilfsmitteln für die Bewältigung des rasant anwachsenden weltweiten Übersetzungsbedarfs konfrontiert Entwickler, Vertrei-

² Im japanischen MÜ-Boom-Jahr 1996 wurden ca. 1 500 000 MÜ-Softwarepakete verkauft (AAMT 1996: 21).

ber und potentielle Anwender von maschineller Übersetzungssoftware einerseits mit der Frage, inwieweit elektronisch bzw. mit elektronischer Unterstützung hergestellter Übersetzungsooutput den Anforderungen des Marktes gerecht werden kann. Auf der anderen Seite sieht man sich zunehmend mit dem Mangel an geeigneten Methoden zur objektiven Beurteilung der Qualität von Übersetzungsooutput³ sowie der Effizienz von maschinellen Übersetzungssystemen im Hinblick auf Kosten und Zeitersparnis – verglichen mit dem ausschließlichen Einsatz von Humanübersetzern – gegenübergestellt. Eine Aufgabe der Forschung muß es daher sein, neue Strategien für die Evaluation von maschinellen Übersetzungssystemen zu entwickeln, zu systematisieren und zu vereinheitlichen. Angesichts der häufigen Fehleinschätzungen und Mißverständnisse im Zusammenhang mit Aspekten der maschinellen Übersetzung besteht eine weitere wichtige Aufgabe darin, die öffentlichen Diskussionen mit Informationen darüber zu bereichern, was solche Systeme gegenwärtig leisten können und was sie möglicherweise in Zukunft leisten werden, aber auch, wo ihre Grenzen liegen. Der vorliegende Beitrag beschäftigt sich mit Problemen des qualitativen Outputs gängiger maschineller Übersetzungssysteme für die Übersetzung geschriebener Texte aus dem Japanischen. Dabei stellt sich die Frage, welche qualitativen Maßstäbe an die Bewertung des Outputs angelegt werden können und welche Bestrebungen in Japan auf dem Gebiet der Forschung und Entwicklung unternommen werden, um den Output maschineller Übersetzungssysteme qualitativ zu verbessern.

MÖGLICHKEITEN UND GRENZEN DES PRAKTISCHEN EINSATZES MASCHINELLER ÜBERSETZUNGEN

Neben rein technischen Beschränkungen für den Einsatz von maschinellen Übersetzungssystemen, die hauptsächlich Speicherkapazitäten und Prozessorgeschwindigkeiten der benutzten Computer betreffen, und der jeweils sprachenpaarabhängigen linguistischen Komponente der maschinellen Sprachverarbeitung, um die es im folgenden gehen wird, bestim-

³ Dies gilt im übrigen nicht nur für maschinell erstellte Übersetzungen, sondern auch für die qualitative Bewertung von Humanübersetzungen. Im Bereich der Übersetzungswissenschaft stand die Frage nach geeigneten Methoden zur Bewertung von Übersetzungen seit Bestehen der Disziplin im Mittelpunkt der wissenschaftlichen Diskussion und führte in den vergangenen Jahrzehnten zur Etablierung verschiedener übersetzungswissenschaftlicher „Schulen“. Einen Überblick gibt z. B. KOLLER (1997).

men viele weitere Faktoren die Möglichkeiten und Grenzen eines praktischen Einsatzes:

- Nicht alle Textsorten sind gleichermaßen für die maschinelle Übersetzung geeignet. Während ein „zufriedenstellender“ Output für den gesamten literarischen Bereich, persönliche Korrespondenz, Fragebögen sowie appellative Texte (z. B. politische Reden und Werbung) auch in Zukunft ausgeschlossen wird und die Ergebnisse im sonstigen alltagssprachlichen Bereich ebenfalls verhältnismäßig dürftig ausfallen, ist man sich darüber einig, daß der Einsatz für Fachtexte, Gebrauchsanleitungen und Handbücher o. ä. ökonomisch und qualitativ sinnvoll ist (z. B. NIRENBURG *et al.* 1992: 8–10, LEÓN und AYMERICH 1997: 6–7).
- Der Einsatz von maschineller Übersetzung ist nur für große Textmengen geeignet.
- Ein „zufriedenstellender“ Output kann nur durch ein gewisses Maß an Prä- bzw. Postedition erzielt werden. Hinzu kommt, daß die meisten maschinellen Übersetzungssysteme interaktiv arbeiten und der Anwender während des Übersetzungsprozesses Entscheidungen treffen muß (z. B. HUTCHINS 1986: 178, NIRENBURG *et al.* 1992: 36, FUJI 1996: 20, LEÓN und AYMERICH 1997: 26–27, NARITA 1997: 54–62). Damit liegt auf der Hand, daß für Benutzer von maschinellen Übersetzungssystemen Kenntnisse beider Sprachen von großem Vorteil sind.
- Kurze Sätze werden mit weniger Fehlern übersetzt als lange Sätze, da kurze Sätze in der Regel weniger syntaktische Ambiguitäten aufweisen (z. B. NAGAO und MAKINO 1995: 54–55, FUJI 1996: 18).
- Die Qualität des Outputs maschineller Übersetzungssysteme hängt in hohem Maße vom Umfang der mitgelieferten Wörterbücher ab. Die Basiswörterbücher enthalten je nach System zwischen zehner- und mehreren hunderttausend Worteinträgen.⁴ Der Benutzer hat fast immer die Möglichkeit, ein Benutzerwörterbuch (*user dictionary*) nach seinen individuellen Bedürfnissen selbst aufzubauen. Die meisten Unternehmen stellen zusätzlich Fachwörterbücher für verschiedene Bereiche zur Verfügung (z. B. NAGAO und MAKINO 1995: 71–82, KOIZUMI 1996: 14–15, LEÓN und AYMERICH 1997: 28–32, NARITA 1997: 71–75).

⁴ Die vom Hersteller angegebenen Größenordnungen für Basiswörterbücher sollten jedoch kritisch hinterfragt werden: Erstens wird in der Regel bei den Zahlenangaben nicht nach den jeweils mitgelieferten Wörterbüchern für Analyse, Transfer und Synthese differenziert, sondern ein Gesamtumfang angegeben. Zweitens wird offengelassen, ob es sich bei den Einträgen im Basiswörterbuch ausschließlich um Allgemeinwortschatz handelt oder ob auch Fachwortschatz integriert ist (SCHWANKE 1991: 40).

- Maschinelle Übersetzungssysteme für strukturell verwandte Sprachen funktionieren „besser“ als für strukturell sehr unterschiedliche Sprachen. Je umfangreicher die syntaktischen Unterschiede zwischen Sprachen sind, desto größer ist die Notwendigkeit einer semantischen Analyse des Ausgangssprachlichen Textes, die in der maschinellen Übersetzung nach wie vor nur ansatzweise geschieht (z.B. NAGAO 1989: viii–ix).
- Aufgrund des unterschiedlichen Analyseaufwands funktionieren maschinelle Übersetzungssysteme, die in die japanische Sprache übersetzen, „besser“ als solche, die aus dem Japanischen in andere – vor allem in europäische – Sprachen übersetzen (z.B. NARITA 1994: 29–31, NARITA 1997: 15–18, 157–159).

Das Hauptproblem jeglicher maschineller Übersetzung ist die allen natürlichen Sprachen immanente Ambiguität sprachlicher Ausdrücke. Dies gilt vor allem für strukturell sehr verschiedene Sprachen wie Japanisch und Englisch bzw. Deutsch. Dabei muß unterschieden werden zwischen Ambiguitäten, die sowohl für Humanübersetzer als auch für die maschinelle Übersetzung mehrdeutig sind, und solchen, die zwar für den Humanübersetzer eindeutig lösbar, für die maschinelle Übersetzung aber mehrdeutig sind.⁵

Beispiel:

Kare wa suki kirai ga hageshii. Buta wa tabenai. – [Er hat stark ausgeprägte Vorlieben und Abneigungen. Schweinefleisch ißt er nicht.]⁶

Was für den Humanübersetzer auf der Hand liegt, ist für ein gängiges maschinelles Übersetzungssystem unmöglich: Es kann sich bei der Interpretation von *Buta wa tabenai* nicht auf den vorangegangenen Satz stützen. Gängige kommerzielle Systeme verarbeiten einen Text, der aus mehreren Sätzen besteht, nicht auf einmal, sondern separieren den Text zunächst in einzelne Sätze. Dieser Prozeß wird *sentence cutout* (jap.: *ichi-*

⁵ Falls nicht anders gekennzeichnet, stammen die angeführten Satzbeispiele und Analysen aus eigenen Testreihen mit der japanischen Software „PC-Transfer/je“ (Version 3.0) der Firma Nova.

⁶ Hier und im folgenden in eckigen Klammern gesetzte deutsche Übersetzungen stammen von mir und sind als Orientierungshilfe gedacht, d.h. sie sind nicht von einem maschinellen Übersetzungssystem übersetzt. Kursiv gesetzte englische Übersetzungen im Zusammenhang mit Beispielsätzen sind i. d. R. aus dem praktischen Betrieb eines maschinellen Übersetzungssystems hervorgegangen und werden nicht mehr ins Deutsche übersetzt.

bun kiridashi) genannt.⁷ Erst dann werden alle Sätze unabhängig voneinander übersetzt (YOSHIMI 1997: 7).

Eine weitere grundlegende Einschränkung für die maschinelle Übersetzung besteht in der Organisation von Wissensquellen. Zur Durchführung einer Übersetzung benötigt man linguistisches Wissen über Ausgangs- und Zielsprache, Fachkenntnisse über das behandelte Thema und nicht-linguistisches Wissen, das in der Fachliteratur als Common sense bzw. Weltwissen bezeichnet wird. Von einem Humanübersetzer wird in der Regel verlangt, über diese Wissensquellen zu verfügen, auch wenn dies in der Praxis längst nicht immer der Fall ist. Ein maschinelles Übersetzungssystem enthält demgegenüber aber nur Teile des linguistischen Wissens und keine der anderen genannten Fähigkeiten; es muß daher die Bedeutung eines Satzes ausschließlich auf der Basis von lexikalischem, syntaktischem und einem Minimum an semantischem Wissen erschließen (SHIRAI 1997: 10–11). Die technische Repräsentation von Fach- und Weltwissen fällt in den Forschungsbereich der künstlichen Intelligenz und der Kognitionsforschung. Bekanntermaßen steckt die Entdeckung der Prozesse im menschlichen Gehirn noch in den Kinderschuhen; zum einen ist es noch völlig unklar, in welcher Form Weltwissen ausgedrückt wird, geschweige denn in welcher Form es per Computer simuliert werden könnte. Zum anderen enthält das Lexikon eines kommerziellen maschinellen Übersetzungssystems bereits viele tausend Einträge; das Hinzufügen von Weltwissen zu jedem einzelnen Eintrag – beispielsweise in Form von semantischen Netzen – wäre in quantitativer Hinsicht viel zu umfangreich und würde enorme Mengen an Arbeitskraft und Entwicklungszeit beanspruchen. Es muß somit betont werden, daß maschinelle Übersetzungssysteme über die Differenzierungsmöglichkeiten eines Humanübersetzers im Hinblick darauf, welche Interpretation bzw. Übersetzung für einen Ausgangssprachlichen Ausdruck qualitativ am besten geeignet ist, hauptsächlich aus zwei Gründen nicht verfügen:

- Humanübersetzer können aus dem Kontext gewonnene Hinweise als Evaluationskriterien für die Interpretation verwenden, maschinelle Übersetzungssysteme nicht.

⁷ Aufgrund des Prinzips des *sentence cutout* kommt es bei der maschinellen Analyse dieses Satzes zu lexikalischen, syntaktischen und semantischen Ambiguitäten: Ist z. B. *buta wa* [Schwein(efleisch)] Subjekt oder Objekt von *tabenai* [nicht essen]? Welches Pronomen wird bei der maschinellen Übersetzung ins Englische bzw. Deutsche für das (nicht explizit genannte) Subjekt gewählt? Bedeutet *buta* in diesem Satz das Tier (*pig/Schwein*) oder das Fleisch (*pork/Schweinefleisch*)? Wie sind die Numerus und Definitheit betreffenden Probleme zu lösen? Diese Mehrdeutigkeiten sind nicht unabhängig voneinander zu betrachten, sondern stehen in enger Beziehung zueinander.

- Im Gegensatz zu maschinellen Systemen können Humanübersetzer als Evaluationskriterien für die Interpretation nicht nur linguistisches Wissen, sondern auch nicht-linguistisches Wissen wie Fach- und Weltwissen verwenden.

PROBLEME DER MASCHINELLEN ÜBERSETZUNG AUS DEM JAPANISCHEN

Die heute in Japan gängigen maschinellen Übersetzungssysteme durchlaufen in der Regel die für sogenannte Transfersysteme⁸ typischen drei Verarbeitungsschritte: Analyse der Ausgangssprache, Transfer zwischen Ausgangs- und Zielsprache, Generierung der Zielsprache. Die Tiefe der Analyse – und somit auch die Qualität des Übersetzungsausputs – hängt davon ab, welche Analyseschritte durchgeführt werden. Die Probleme der maschinellen Übersetzung aus dem Japanischen werden anhand dieser Schritte (morphologische, syntaktische und semantische Analyse) dargestellt.

LEXIKALISCH-MORPHOLOGISCHE ASPEKTE

Die morphologische Analyse beginnt mit der Zerlegung eines Satzes in seine Worteinheiten. Anders als z.B. im Englischen oder Deutschen werden die Wörter im geschriebenen Japanisch nicht durch Leerstellen in individuelle Einheiten zerlegt, so daß die morphologische Analyse des Japanischen vergleichsweise komplex ausfällt. Um zu ermitteln, wo ein Wort beginnt und endet, muß eine Folge aneinandergereihter Schriftzeichen vorläufig extrahiert und die An- oder Abwesenheit dieser Folge im Lexikon überprüft sowie bei Anwesenheit ihre Wortklasse bestimmt werden (NAGAO 1989: 71–72). Je mehr Kanji verwendet werden, desto weniger Ambiguitäten fallen bei der morphologischen Analyse an (Beispiele aus NAGAO und MAKINO 1995: 45, SHU 1997: 23):

Beispiel:

Kuruma de / hako o / hakobu – [Eine Schachtel mit dem Auto transportieren.]

⁸ Übersichten über die unterschiedlichen Methoden maschineller Übersetzung finden sich etwa bei HUTCHINS (1986: 174–188), LEHRBERGER und BOURBEAU (1988: 11–38), HUTCHINS und SOMERS (1992: 69–76), NIRENBURG *et al.* (1992: 13–18). Die kommerziellen japanischen maschinellen Übersetzungssysteme gehören ohne Ausnahme zu den Transfersystemen. Diese sind für die bilinguale maschinelle Übersetzung geeignet, während andere Methoden (z.B. Interlinguasysteme) eher für multilinguale maschinelle Übersetzung entwickelt werden.

Kuruma de wa / ko o / hakobu – [Kinder mit dem Auto transportieren.]

Kuru made / hako o / hakobu – [Eine Schachtel bis zum Kommen transportieren/tragen.]

Kuru made wa / ko o / hakobu – [Kinder bis zum Kommen transportieren/tragen.]

Beispiel:

Niwa ni wa niwa niwatori ga iru. – [Im Garten sind zwei Hühner.]

Sumomo mo momo mo momo no uchi. – [Sowohl (gelbe) Pflaumen als auch Pfirsiche gehören zu den Pfirsichgewächsen.]

Schreibt man diese Beispielsätze ausschließlich in Hiragana, wären für die morphologische Analyse zahlreiche Interpretationsmöglichkeiten gegeben.

Zu den lexikalisch-morphologischen Besonderheiten der japanischen Sprache, die bei einer maschinellen Übersetzung ins Englische bzw. Deutsche problematisch sind, gehören das Fehlen von Definitheit, Numerus und Geschlechtsangaben sowie andere Tempuskategorien als im Englischen/Deutschen.

Beispiel:

Watashi wa gakusei desu. – [Ich bin Student/Studentin.]

Der gleiche Satz kann sich ohne Veränderung der Verbform (z.B. durch Ergänzung einer adverbialen Bestimmung) auf ein Ereignis in der Zukunft beziehen:

(Rainen) watashi wa gakusei desu. – [(Im nächsten Jahr) werde ich Student/Studentin sein.]

Das Fehlen von Definitheits- und Tempusinformation im Japanischen stellt ein Hauptproblem der maschinellen Übersetzung ins Englische bzw. Deutsche dar, da in diesen Sprachen zwischen bestimmten und unbestimmten Artikeln sowie Singular und Plural unterschieden wird.⁹ Singular und Plural werden im Japanischen auf der lexikalischen Ebene

⁹ Ausnahmen, in denen auch japanische Nominalphrasen Definitheitsinformation enthalten, bilden einige attributive Demonstrativa (*kono, sono, ano* und das Fragewort *dono* [dies, das, jenes/das, welch(es)]) sowie bestimmte Adjektive und Genitivkonstruktionen (z.B. *ouaji, tsugi no, kondo no* [der/die/das gleiche/selbe, der/die/das nächste, der/die/das kommende/nächste/letzte] u.ä.) (SIEGEL 1997: 146–147).

i. d. R. nicht unterschieden (FUJINAMI und NANZ 1997: 3–4, SIEGEL 1997: 145).¹⁰

Beispiel:

hon – [ein Buch, das Buch, Bücher, die Bücher]

Lexika für maschinelle Übersetzungssysteme unterscheiden sich stark von „normalen“ Lexika. Als zweite Form von Wissensbasen neben den Grammatiken stellen solche Lexika folgende Informationen bereit:

- (1) Linguistische Phänomene für jedes Wort (z.B. Flexionsformen, Satzpattern etc.)
- (2) Übersetzungsrelationen
- (3) Informationen zur Generalisierung linguistischer Phänomene: Um jedes Wort den in der Grammatik für maschinelle Übersetzungssysteme aufgeführten allgemeinen Regeln zuordnen zu können, ist es notwendig, die Klassifizierung von Wörtern im Lexikon zu beschreiben (Wortklassen, semantische Kategorien etc.)

Die auffälligsten Unterschiede zwischen Lexika für maschinelle Übersetzungssysteme und für Humanübersetzer bestehen zum einen in der engen Korrespondenz des maschinell lesbaren Lexikons mit der maschinell lesbaren Grammatik und zum zweiten in der Art und Weise der Beschreibungen. Kein Lexikon für maschinelle Übersetzungssysteme kann eine Verbindung zwischen einzelnen Wörtern und Grammatikregeln herstellen, ohne genaue Übereinstimmung der Klassifizierung des Wortes in Wortklassen und semantische Kategorien mit dem verallgemeinerten Klassifizierungsschema in der Grammatik. Humanübersetzer können selbst bei Abweichungen der Klassifizierungsmaßstäbe Übereinstimmungen feststellen, maschinelle Übersetzungssysteme sind in dieser Hinsicht inflexibel. Die Rolle eines Wortes im Satz sowie die Klassifizie-

¹⁰ Während der syntaktischen Analyse können allerdings Numerale Aufschluß über Tempus- und Numerusinformation geben (z.B. *ichijikan, sanjikan/hitori, gonin* [eine Stunde, drei Stunden/ eine Person, fünf Personen] etc.). Und in einigen Substantiven (z.B. *mina* [alle(s)]) ist der Wert „Plural“ in der lexikalischen Information bereits enthalten (SIEGEL 1997: 149). Eine ausführliche Diskussion über die Problematik Definitheit und Numerus fand im Rahmen des VERBMOBIL-Forschungsprojekts an der Universität Bielefeld statt (vgl. SIEGEL 1994, SIEGEL 1997: 145–153). VERBMOBIL ist ein automatisches Dolmetschsystem für die Sprachen Deutsch/Englisch/Japanisch. Die Forschungen beschäftigen sich daher mit der Problematik gesprochener Sprache, die sich in vielerlei Hinsicht vom Gegenstand der maschinellen Übersetzung geschriebener Texte unterscheidet.

rung des Wortes erfolgt in Lexika für Humanübersetzer meist ebenfalls in Worten bzw. beispielhaft veranschaulicht. Ein Lexikon für die maschinelle Übersetzung erfordert demgegenüber explizite und formale Beschreibungen (KOIZUMI 1996: 10–11).

Als Eintrag wird registriert, was bei der Übersetzung als minimales Kompositionselement eines Satzes behandelt wird. Die Registrierung von semantischen Einheiten (*imateki na matomari*) wie z. B. *ka mo shirenai* [vielleicht, wohl] kann sowohl Zeit bei der maschinellen Verarbeitung einsparen als auch zur Vermeidung von Fehlern bei der morphologischen Analyse beitragen. Weitere Beispiele sind *zaru o enai* [nicht umhinkönnen, zu ...], *ta koto ga aru* [schon (einmal) ... gemacht haben], *te wa naranai* [nicht ... tun dürfen]. Solche semantischen Einheiten werden in der maschinellen Übersetzung zumeist als Postpositionen oder als Hilfsverben klassifiziert und entsprechen somit nicht den schulgrammatischen Verbindungsstrukturen, sondern eher den tatsächlichen Anwendungen. Ebenso werden (Substantiv-)Komposita wie etwa *kikai hon'yaku* [maschinelle Übersetzung] behandelt: Verglichen mit Lexika für Humanübersetzer müssen in Lexika für die maschinelle Übersetzung eine größere Menge an Komposita und Idiomen aufgeführt werden; problematisch dabei ist, daß deren Anzahl fast grenzenlos ist. Während ein Lexikon für die maschinelle Übersetzung sämtliche erforderlichen Komposita einzeln enthalten muß (seien sie auch noch so ähnlich), reicht in einem Lexikon für Humanübersetzer im Prinzip ein einziges Beispiel (z. B. *vitamin-enriched food*), um die Anwendung ähnlicher Ausdrücke (wie *mineral-enriched food*, *calcium-enriched food* etc.) ableiten zu können (NAGAO 1989: 62–64).¹¹

„Selektionsbeschränkung“ nennt man die Beschreibung von Informationen im Lexikon, die darüber Aufschluß geben, welcher Wortinhalt in dem jeweiligen syntaktischen Element (Subjekt, Objekt, Komplement, präpositionales Objekt etc.) im Hinblick auf Kategorien wie Verb, Substantiv, Adjektiv gerade vorkommt. Selektionsbeschränkungen sind für die Auswahl eines geeigneten Wortes oder die Auflösung bestimmter Ambiguitäten wichtig.

Beispiel:

[<*hito ga*> <*densenbyō ni*> <*kansen suru*>]

¹¹ Alle in Japan derzeit im Handel erhältlichen maschinellen Übersetzungssysteme arbeiten nach dem Prinzip der „Kompositionalität“ (vgl. NAGAO 1989: 62–64): Bei der morphologischen Analyse werden Zeichenketten prinzipiell bis in die kleinste im Lexikon auffindbare Einheit zerlegt, es sei denn, längere Zeichenketten sind dort als idiomatische Wendungen bzw. als semantische Einheiten registriert.

[<man> <with contagious disease> <is infected>]

= Beschreibung einer der verbalen Bedeutungen von *to catch* in einem Lexikon für Humanübersetzer

Lexika für maschinelle Übersetzungssysteme beschreiben sehr ähnliche Selektionsbeschränkungen. Der Hauptunterschied besteht darin, daß bei Anwendung einer semantischen Kategorie in der Selektionsbeschränkung ein semantischer Klassifikationscode für das Wort, das auf diese semantische Kategorie paßt, beschrieben wird. In einem Lexikon für Humanübersetzer fehlen semantische Kategorien wie <*densenbyō*> [ansteckende Krankheit] bei der Beschreibung der in Frage kommenden Substantive. Wenn ein solches Lexikon etwa für die Übersetzung von *measles* das japanische Wort *mashin* [Masern] vorgibt, verstehen Humanübersetzer aufgrund von außerlinguistischem Wissen, daß Masern eine ansteckende Krankheit sind. Kein maschinelles Übersetzungssystem kann jedoch beurteilen, daß der Ausdruck *catch the measles* die im Beispiel angeführte Selektionsbeschränkung befriedigt – es sei denn, in dem Worteintrag *measles* würde die semantische Kategorie <*densenbyō*> beschrieben. Um *catch the measles* mit *mashin ni kakaru* übersetzen zu können, besteht zusätzlich die Möglichkeit der Eintragung als Idiom. Entsprechende Idiome müßten dann aber für alle ansteckenden Krankheiten (*diphtheria*, *mumps* etc.) in das Lexikon für die maschinelle Übersetzung eingetragen werden.

Solche Lexika beschreiben viele linguistische Phänomene, die nicht zwangsläufig auch in Lexika für Humanübersetzer eingetragen sind, wie z. B. Zählwörter. In einigen rein japanischen und den meisten japanisch-englischen oder japanisch-deutschen Lexika für Humanübersetzer finden sich bei Einträgen wie *usagi* [Hase] oder *karasu* [Krähe] keine Hinweise auf ihre Zählweise. Entsprechend findet sich aber bei einem Eintrag wie *wa* [= Numeralklassifikator für Vögel und Hasen] die Erklärung „wird zur Zählung von Vögeln oder Hasen an Numerales angehängt“ o. ä. Eine korrekte Anwendung solcher Hinweise bei der Übersetzung von *sieben Krähen* in *nana-wa no karasu* ist nicht nur für maschinelle Übersetzungssysteme unmöglich, sondern stellt auch Humanübersetzer oftmals vor Probleme. In Lexika für die maschinelle Verarbeitung müssen die Anwendungsarten von Numeralklassifikatoren deutlich beschrieben werden (KOIZUMI 1996: 11–12).

Bei der Übersetzung von Sprachen mit sehr unterschiedlichen Satzstrukturen wie Japanisch und Englisch bzw. Deutsch ist es in vielen Fällen besser, Substantive als Verben zu übersetzen oder umgekehrt. Die meisten Lexika für Humanübersetzer beschreiben aber nur Übersetzungen von korrespondierenden Wortklassen (z. B. *complete* – *oeru* [vollenden],

completion – *kanryō* [Vollendung] etc.). Während eine flexible Übertragung von Wortklassen für Humanübersetzer kein Problem darstellt, betrachtet das maschinelle Übersetzungssystem beide Sprachen als Fremdsprachen; Lexika für den maschinellen Gebrauch müssen daher Übersetzungen für verschiedene Wortklassen bereitstellen, da etwa eine Übersetzung von *complete* mit *kanryō* bzw. von *oeru* mit *completion* ansonsten nicht möglich wäre. Solche Lexika beschreiben aber nicht nur Derivativa, bei denen die Bedeutungen gleich, die Wortklassen aber unterschiedlich sind, sondern stellen auch Übersetzungen für Derivativa bereit, die Negation, Potentialis etc. ausdrücken.

Beispiel:

Für *applicable* als eine mögliche Übersetzung von *tekiyō* [anwenden, anwendbar, Verwendung u.ä.] wird der Lexikoneintrag für die maschinelle Übersetzung durch die Information <*tenka imi*: +*kanō*> [Zusatzbedeutung: +Potentialis] ergänzt.

Zur Übersetzung von Ausdrücken wie *tekiyō dekiru*, *tekiyō shiuru*, *tekiyō ga kanō da*, *tekiyō suru koto ga dekiru* als *applicable* sind erstens eine Analyserregel zur Interpretation des jeweiligen Ausdrucks als *tekiyō* <+*kanō*>, zweitens ein Lexikon mit derivativen Informationen wie oben beschrieben sowie drittens eine generative Regel (z.B. „Wenn es ein Übersetzungswort mit <+*kanō*> gibt, verwende dieses Wort für die Übersetzung.“) notwendig. Im Japanischen häufig vorkommende Derivativa enthalten nicht nur Präfixe wie *hi-*, *fu-*, *mi-* und Suffixe wie *-sei*, *-teki*, *-ka*, sondern auch Substantivierungen von Verben wie *tatakai* [der Kampf], *warikomii* [das Hineindrängen/Einmischen] etc. (KOIZUMI 1996: 12–13).

SYNTAKTISCHE ASPEKTE

Nach Abschluß der morphologischen Analyse folgt die Analyse der syntaktischen Struktur der einzelnen Sätze. In maschinellen Übersetzungssystemen werden für die syntaktische Analyse viele hundert Grammatikregeln verwendet.¹² Das Hauptproblem der syntaktischen Analyse ist die

¹² Die meisten heute gängigen maschinellen Übersetzungssysteme benötigen drei Grammatiken: eine Analysegrammatik zum Verständnis der Strukturen der Ausgangssprache (*kaiseki bunpō*), eine Transfergrammatik zur Konvertierung der ausgangssprachlichen Strukturen in die Zielsprache (*henkan bunpō*) und eine Synthesegrammatik zur Generierung der Zielsprache in die der zielsprachlichen Grammatik entsprechenden Strukturen (*seisei bunpō*). Grammatische Regelwerke für maschinelle Übersetzungssysteme sind wesentlich präziser formuliert als Schulgrammatiken. Japanische maschinelle Überset-

Auflösung von strukturellen Mehrdeutigkeiten (NAGAO 1989: 71–74). Viele strukturelle Ambiguitäten können weder mit Hilfe der syntaktischen noch mit Hilfe der semantischen Analyse aufgelöst werden, wie z.B. viele Sätze mit mehrdeutigen Bezügen (Beispiele aus NAGAO und MAKINO 1995: 34, SHU 1997: 24):

Beispiel:

Watashi wa hashi de nete iru hito o mita. – [Auf der Brücke (= als ich auf der Brücke stand / von der Brücke aus o. ä.) sah ich eine Person, die schlief. / Ich sah eine Person, die auf der Brücke schlief.]

Beispiel:

Keisanki de keisan shita kekka o kenshō suru. – [Die Resultate, die man mit dem Computer errechnet hat, verifizieren. / Die errechneten Resultate mit dem Computer verifizieren.]

Als weitere für die maschinelle Übersetzung aus dem Japanischen ins Englische bzw. Deutsche bislang unlösbare Aufgabe hat sich die Übersetzung parataktischer Phrasen (*heiretsuku*) erwiesen (Beispiele aus NAGAO 1989: 93, FUJI 1996: 18):

zungssysteme der ersten Generation basierten auf einfachen japanischen Schulgrammatiken. Mit der Zeit wurden immer neue Grammatikregeln hinzugefügt, die zunächst als Ausnahmen von den zugrunde gelegten Grammatikregeln formuliert wurden. Mit anwachsender Anzahl dieser Ausnahmeregeln entstanden zunehmend Widersprüche. Das Scheitern dieser ausschließlich grammatikorientierten Methode für die maschinelle Übersetzung führte in Japan zur Entwicklung einer beispielsatzorientierten Herangehensweise. Als Grundlage dient ein sprachlicher Korpus, der eine große Anzahl von Paaren bereits übersetzter Sätze enthält. Bei Eingabe eines beliebigen Satzes in das maschinelle Übersetzungssystem werden Übersetzungspaare aus diesem Korpus durchsucht, die dem Inputsatz strukturell ähnlich sind; dann wird die geeignetste Übersetzung ausgewählt und die alten Übersetzungswörter werden durch die dem Inputsatz entsprechenden Übersetzungswörter ersetzt, z.B. *Atarashii shamei ga yūkō ni naru.* (*The change of the company's name takes effect.*) – *Atarashii kōmei ga yūkō ni natta.* (*The change of the school's name took effect.*). Auf den ersten Blick schien die Beispielsatz-Methode den Einsatz von Grammatikregeln überflüssig zu machen. Aber auch die für die Beschreibung der zu ersetzenden Ausdrücke notwendigen Regeln erfordern natürlich die Anwendung grammatischer Kategorien (für das o. a. Beispiel etwa die Regel, daß xxx in dem Ausdruck „the change of xxx's ...“ durch bestimmte Nomen ersetzt werden kann). Es gibt bis heute keine Methode für die maschinelle Übersetzung, bei der auf die Verwendung grammatischer Elemente verzichtet werden könnte (FUJI 1996: 17–18).

Beispiel:

Tarō to Hanako no inu – (Tarō to Hanako) no inu / Tarō to (Hanako no inu)
[Der Hund von (Tarō und Hanako) / Tarō und (Hanakos Hund)]

Beispiel:

*denshi kairo no (denatsu to denryū) – the voltage and the current of the
electronic circuit*

aber:

*(denshi kairo no zatsuen) to zentai shisutemu – the electronic circuit's noise
and the entire system*

Syntaktische Phänomene, in denen sich das Japanische vom Englischen bzw. Deutschen unterscheidet und die bei der maschinellen Übersetzung Probleme bereiten, sind z.B. die verhältnismäßig freie Wortreihenfolge (ähnlich dem Russischen), die Verwendung von Postpositionen zum Anzeigen grammatischer Funktionen,¹³ das Fehlen von Relativpronomina, Topikmarkierungen, Formalsubstantive, Leidenspassiv und Potentialis (NAGAO 1989: 93–104, NARITA 1994: 35–38, FUJINAMI und NANZ 1997: 4–7, 37–39). Das Fehlen von Relativpronomina kann vor allem in Kombination mit Nullpronomina zu Ambiguitäten führen:

Beispiel:

*Hanako ga katta kōhī o nomimasu. – Hanako drinks the coffee that she
bought. / Hanako drinks the coffee that <I, he, Tarō etc.> bought. / <I, he,
Tarō etc.> drink(s) the coffee that Hanako bought.*

Die syntaktische Analyse erkennt entweder (1) *Hanako ga* als Subjekt des Hauptsatzes und entscheidet, (a) daß das Subjekt der Relativkonstruktion ebenfalls *Hanako ga* ist, oder (b) daß das Subjekt der Relativkonstruktion unbekannt ist; außerdem kann (2) *Hanako ga* als Subjekt der Relativkonstruktion erkannt und entschieden werden, daß das Hauptsatz-Subjekt fehlt.

Die im Japanischen häufig vorkommenden doppelten Verneinungen können besondere Probleme für die maschinelle Übersetzung verursachen:

¹³ Erhebliche Probleme bereiten die Postpositionen *ni* und *de* mit ihren zahlreichen grammatischen Funktionen sowie die topikmarkierende Postposition *wa* und die Austauschbarkeit der Postpositionen *o* und *ga* in vielen Fällen. Die entsprechenden Regeln werden in Tabellenform dargestellt und sind sehr umfangreich (NAGAO und MAKINO 1995: 41–45).

Beispiel:

Mattaku imi no nai bunshō to wa ienai. – It cannot be said that the sentence has no meaning at all. / It absolutely cannot be said that the sentence has no meaning.

Japanische Sätze, die aufgrund der freien Wortstellung unterschiedliche syntaktische Oberflächen, aber die gleiche Grundbedeutung besitzen, müssen in der syntaktischen Analyse gleiche Ergebnisse erzielen, auch wenn die Haltung des Sprechers und die Betonungen innerhalb der Sätze variieren (Beispiel aus NAGAO 1989: 68–69):

Beispiel:

Watashi wa kinō Shinkansen de Tōkyō e itta. / *Kinō watashi wa Shinkansen de Tōkyō e itta.* / *Watashi wa kinō Tōkyō e Shinkansen de itta.* – [Gestern fuhr ich mit dem Shinkansen nach Tōkyō.]

Anhand der Verarbeitung der japanischen Wortklassen (*hinshi*) wird in der Folge veranschaulicht, warum die Regeln der japanischen Schulgrammatik für die morphologische und syntaktische Analyse der maschinellen Übersetzung nicht ausreichend sind.¹⁴ Bei der morphologischen Analyse wird – wie bereits erläutert – jeder Satz in seine bedeutungstragenden Einheiten zerlegt. Jedes Morphem ist im Lexikon registriert und mit der jeweiligen Wortklasse versehen. Bei der maschinellen Analyse des folgenden Satzbeispiels wird das System mit Interpretationsmöglichkeiten konfrontiert, die für einen Humanübersetzer undenkbar wären:

Beispiel:

Kanojo wa gaikoku ni ikitai sō da. – [Ich habe gehört, daß sie ins Ausland gehen will.]

¹⁴ Hiermit zusammenhängend ist auch die Verarbeitung der Flexionsstufen zu betrachten. Die maschinelle Übersetzung verwendet weder genau die in der japanischen Schulgrammatik genannten Wortklassen und Flexionsstufen noch völlig andere. Im Falle der Verben ist es z.B. von Vorteil, *mizenkei* [Verbal-Indefinitform] und *ren'yōkei* [Konjunkionalform] in jeweils zwei Kategorien zu unterteilen. Statt sechs werden auf diese Weise acht Flexionsstufen verwendet. Wenn außerdem die in einer Flexionstabelle aufgelisteten unveränderbaren Bestandteile als Stammformen der oberen und unteren einstufigen Verben betrachtet werden, fallen sie insgesamt in die gleiche Flexionskategorie. Grammatiken für die maschinelle Übersetzung werden mit diesen Kunstgriffen einerseits detaillierter als Schulgrammatiken, andererseits aber in vielen Fällen auch einfacher (KUMANO 1996: 18).

Entsprechend der morphologischen Analyse wären laut Lexikon folgende Zerlegungen möglich:

- (1) *iki* (Verb) / *tai* (Hilfsverb) / *sō da* (Hilfsverb)
- (2) *iki* (Verb) / *tai* (Hilfsverb) / *sō* (Adverb) / *da* (Hilfsverb)
- (3) *iki* (Verb) / *ta* (Hilfsverb) / *i* (Verb) / *sō da* (Hilfsverb)
- (4) *iki* (Verb) / *ta* (Hilfsverb) / *i* (Verb) / *sō* (Adverb) / *da* (Hilfsverb)

Mit Hilfe der syntaktischen Analyse beurteilt das maschinelle Übersetzungssystem als erstes (3) und (4) als falsch: das Hilfsverb *ta* wird nicht direkt an die Konjunkionalform *iki* des Verbs *iku* [gehen/fahren] angefügt. In (1) und (2) wird *iki* als *ren'yōkei* [Konjunkionalform] für das Hilfsverb *tai* interpretiert. Bei (2) wird jedoch erkannt, daß *tai shūshikei* [Finalform] oder *rentaikei* [Attributivform] ist und kein Adverb direkt an die *shūshikei* anschließen kann. Wenn das System nun davon ausgeht, daß es sich um die *rentaikei* handelt, so wird auch diese Interpretation als falsch erkannt, weil danach kein Substantiv mit adnominaler Bestimmung folgt. Aufgrund solcher Interpretationsprobleme ist es für die Analyse der Ausgangssprache absolut notwendig, erkennen zu können, ob eine beliebige Flexionsform eines beliebigen Morphems sich mit einem anderen Morphem verbinden kann oder nicht. Eine einfache Methode für die maschinelle Übersetzung besteht darin, alle möglichen Kombinationen in einer Tabelle aufzulisten. Diese „Tabelle grammatischer Verknüpfungen“ (*chokusetsu tēburu*) ist für die morphologische Analyse des Japanischen besonders geeignet (KUMANO 1996: 19).¹⁵

Beispiel:

Kyō honnin kara kikimashita. – [Ich habe es heute direkt von ihm gehört.]

Entsprechend der morphologischen Analyse wären laut Lexikon für die erste Satzhälfte folgende Zerlegungen möglich:

¹⁵ Für eine Tabelle grammatischer Verknüpfungen benötigt man so viele Zeilen, wie unterschiedliche Flexionsformen der jeweiligen Flexionswörter existieren. Da die *shūshikei* der fünfstufigen Verben unabhängig von den Flexionszeilen gleiche Verbindungen eingeht, kann sie zu einer Kategorie zusammengefaßt werden. Durchschnittlich sind für eine Tabelle ca. zweihundert Zeilen notwendig; ca. einhundert Spalten werden für die verschiedenen Arten von Postpositionen, Hilfsverben, Formalsubstantiven, modalen Hilfsverben etc. benötigt. Solche Tabellen sind wesentlich detaillierter als schulgrammatische Vorgaben. Ein Tabellenbeispiel findet man bei KUMANO (1996: 19).

- (1) *kyō / honnin / kara / kikimashita*
- (2) *kyō / hon / hito / kara / kikimashita*
- (3) *ima / Nihonjin / kara / kikimashita*
- (4) *ima / Nihon / hito / kara / kikimashita*
- (5) *ima / hi / honnin / kara / kikimashita*
- (6) *ima / hi / hon / hito / kara / kikimashita*

(*kikimashita* wird als *kiki + mashi + ta* zerlegt)

Kyō [heute], *ima* [jetzt], *Nihonjin* [Japaner(in)], *Nihon* [Japan], *hi* [Sonne/Tag], *honnin* [er/sie selbst], *hon* [Buch] und *hito* [Mensch] sind Substantive. Aber einige von ihnen unterscheiden sich durch die Rollen, die sie im Satz einnehmen. Substantive werden von direkt anschließenden Postpositionen begleitet und werden dadurch zu grammatischen Kategorien wie z.B. Subjekt oder Objekt (wie etwa das *honnin* in *honnin kara kikimashita*). Das Substantiv *kyō* hingegen wird neben der Verwendung mit Postposition (wie etwa in *kyō wa samui* [Heute ist es kalt.]) auch als Adverb (wie etwa in *kyō kare ni atta* [Heute habe ich ihn getroffen.]) verwendet. Gleiches gilt für die Verwendung von *ima*, aber nicht für *Nihon*, *Nihonjin*, *honnin* etc. Diese Eigenschaft ist für die syntaktische Analyse von großer Bedeutung. Dabei werden Wörter wie *kyō* oder *ima* von „normalen“ Substantiven separiert und als „adverbiales Substantiv“ (*adverb-like noun, fukushiteki meishi*) gesondert behandelt. Anschließend wird anhand der syntaktischen Analyse die richtige Interpretation gesucht und die Versionen (1) und (3) als korrekte japanische Sätze akzeptiert.

Die morphologische Analyse des folgenden Beispielsatzes ist schulgrammatisch umfangreicher als für eine Grammatik für die maschinelle Übersetzung:

Beispiel:

Sore wa hontō kamo shirenai. – [Das ist vielleicht wahr.]

Analyse: *sore* (Demonstrativpronomen, *shiji-daimeishi*) *wa* (Adverbiale Postposition, *fukujoshi*) *hontō* (Stammform eines verbalen Qualitativums, *keiyōdōshi no gokan*) *ka* (Adverbiale Postposition, *fukujoshi*) *mo* (Adverbiale Postposition, *fukujoshi*) *shire* (Verbal-Indefinitform, *dōshi no mizenkei*) *nai* (Finalform eines Hilfsverbs, *jōdōshi no shūshikei*)

Bei der Übersetzung ins Englische würde ein Humanübersetzer *sore wa hontō da* als *it is true* und inklusive der Ergänzung *kamo shirenai* als *it*

may be true übersetzen, wobei die schulgrammatische Klassifizierung von *kamo shirenai* wie oben beschrieben keine Anwendung finden würde. Gleiches geschieht bei der maschinellen Übersetzung dieses Satzes: *kamo shirenai* wird bereits im Lexikon für maschinelle Übersetzungssysteme als Einheit (hier: als *jodōshi*) aufgeführt und bei der morphologischen Analyse als Einheit mit *may* wiedergegeben. Hintereinander stehende Bestandteile eines Satzes, die nach der automatischen Separierung in Wortklassen keine Bedeutung ergeben, werden wieder zusammengesetzt und als Einheit behandelt. Die morphologische Analyse von semantischen Einheiten ist eine programmiertechnisch rationale Methode (KUMANO 1996: 20). Weitere semantische Einheiten, die schulgrammatisch detaillierter analysiert, bei der maschinellen Verarbeitung aber als postpositionale Einheiten betrachtet werden, sind z.B. *ni yotte, o mochiite, no tame ni* u.ä.

SEMANTISCHE UND PRAGMATISCHE ASPEKTE

Bis Mitte der sechziger Jahre wurde in der Forschung vielfach die Ansicht vertreten, eine Analyse von Sätzen sei gänzlich ohne Einbeziehung semantischer Aspekte zu bewältigen. Nach der vernichtenden Kritik im ALPAC-Report¹⁶ gelangte man – wenn auch widerwillig – allmählich zu der Erkenntnis, daß die Anwendung semantischer Methoden für die maschinelle Übersetzung unerlässlich sei (NAGAO 1989: 76). Im Rahmen eines Forschungsprojekts an der Universität Kyōto wurden spezielle semantische Kategorien (sogenannte *semantic primitives*: z.B. *animal, human, plant, inorganic matter, place, time*) zur Eliminierung semantischer Ambiguitäten entwickelt, die heute in fast allen japanischen maschinellen Übersetzungssystemen zum Einsatz kommen. Bei der Analyse einer Folge von Nominalphrasen werden etwa diejenigen Substantive, die mit in Beziehung zueinander stehenden semantischen Kategorien versehen sind, zweckmäßiger zusammen plaziert als diejenigen mit unterschiedlichen Bedeutungsstrukturen.¹⁷ Mit solchen semantischen Kategorien kön-

¹⁶ 1964 berief die National Academy of Sciences in den USA mehrere Gutachter zur Untersuchung von Stand und Perspektiven der maschinellen Übersetzung. Das daraus entstandene Gutachten des Automatic Language Processing Advisory Committee (ALPAC-Report) führte zur Streichung sämtlicher staatlicher Forschungsinvestitionen in den USA und zu einer weltweiten Krise der Forschung im Bereich der maschinellen Übersetzung (ALPAC 1966: 11–24).

¹⁷ Dabei muß zusätzlich die Beschaffenheit des Verhältnisses zwischen den betroffenen Wörtern in Tabellenform explizit gemacht werden (z.B. übergeordnetes / untergeordnetes Konzept, Synonym, Antonym). Ein übergeordnetes Konzept für das Wort *children* wäre *human being*; ein übergeordnetes Konzept für

nen auch die vielen verschiedenen Bedeutungen der japanischen Postpositionen ermittelt werden (NAGAO 1989: 79–82).

Elliptische Phänomene, d.h. das Fehlen von für die Zielsprache obligatorischen Argumentpositionen des Verbs in der Ausgangssprache, sind im Japanischen häufig anzutreffen und ein Hauptproblem für die semantische Analyse.¹⁸ Subjekt und / oder Objekt können im japanischen Satz ausgelassen und ohne Kontextwissen nicht ermittelt werden (z.B. NAGAO 1989: 121, NAGAO und MAKINO 1995: 48–50, FUJINAMI und NANZ 1997: 8, SIEGEL 1997: 72–84). Eine einfache, wenn auch stilistisch selten tolerierbare Lösung ist in einigen Fällen die Bildung von Passivkonstruktionen:

Beispiel:

Tabemono o ataeta. – Food was given.

Beispiel:

Jikken to hikaku shi yoi itchi o mita. – Comparison was made with experiment and good agreement was seen.

Eine weitere für die maschinelle Übersetzung typische semantische Mehrdeutigkeit, die ohne Kontextwissen nicht aufgelöst werden kann, ist die Unterscheidung von Appellativa (*futsū meishi*) und Eigennamen (*koyū meishi*):

Beispiel:

Nihon to Igirisu ryōkoku no shushō ga tomo ni kisha kaiken ni nozonda. – [Die Premierminister Japans und Englands wohnten gemeinsam dem Presseinterview bei.]

(Verwendung von *ryōkoku* als Appellativum = „beide Länder“)

Beispiel:

Ōzumō Hatsubasho wa 12 nichi kara Ryōgoku no Kōkūgikan de hajimaru. – [Das Sumō-Neujahrsturnier beginnt in zwölf Tagen in der Sumō-Halle von Ryōgoku.]

(Verwendung von *Ryōgoku* als Ortsname)

die semantische Kategorie *human being* wäre *animal*. Untergeordnete Konzepte für *children* wären z.B. *infant* oder *student*. Eine Tabelle mit sämtlichen semantischen Kategorien samt ihrer Beziehungsstrukturen wird Thesaurus genannt. Ein Beispiel findet sich in NAGAO (1989: 78).

¹⁸ Für das Dialogdolmetschen von besonderer Relevanz sind Nullpronomina, die im gesprochenen Japanisch sehr häufig vorkommen. Der Thematik wurde im Rahmen des VERBMOBIL-Projekts eine eigene Studie gewidmet (METZING und SIEGEL 1994).

Als ein Grund, warum Humanübersetzer die jeweils richtige Bedeutung von *ryōkoku/Ryōgoku*, die mit identischen Schriftzeichen geschrieben werden, auf Anhieb erfassen können, wird angenommen, daß Humanübersetzer für die Beurteilung „Schlüsselwörter“ finden können und anschließend die syntaktischen Relationen zwischen Schlüsselwörtern und *ryōkoku/Ryōgoku* analysieren. Die Schlüsselwörter, die im ersten Beispiel auf die Bedeutung „beide Länder“ hindeuten, sind *Nihon* [Japan], *Igirisu* [England] und *shushō* [Premierminister]. Die Schlüsselwörter, die im zweiten Beispiel auf den Eigennamen *Ryōgoku* hinweisen, sind *Ōzumō Hatsubasho* [Sumō-Neujahrsturnier] und *Kokugikan* [Sumō-Halle] (YOSHIMI 1997: 9–10).¹⁹ Auch bei der Unterscheidung von Orts- und Personennamen im Japanischen können Humanübersetzer auf Schlüsselwörter referieren (Beispiel aus SHU 1997: 24):

Beispiel:

Watashi wa Murayama de katsute sannen gurai Nihongo no benkyō o shita koto ga aru. – [Ich habe einmal etwa drei Jahre lang in Murayama Japanisch studiert. / Ich heiße Murayama und habe einmal drei Jahre lang Japanisch studiert.]

Watashi wa Murayama de katsute Nihon-koku no seifu sōri daijin o shita koto ga aru. – [Ich heiße Murayama und war einmal Premierminister der japanischen Regierung.]

Im ersten Satz kann weder ein Humanübersetzer noch ein maschinelles Übersetzungssystem entscheiden, ob es sich um eine Person namens Murayama oder um den Ort Murayama in der Präfektur Yamagata handelt. Im zweiten Satz ist es für den Humanübersetzer sofort ersichtlich, daß vom ehemaligen Premierminister Murayama Tomiichi die Rede ist.

Lokale Angaben sind im Englischen bzw. Deutschen manchmal perspektivisch anders als im Japanischen:

Beispiel:

Handoru no mae ni suwatte kudasai. – *Sit down in front of the driving wheel.*

¹⁹ Die letztendliche Entscheidung für die Übersetzung erfordert aber nicht nur das Ausmachen eines oder mehrerer Schlüsselwörter, sondern auch die Analyse der syntaktischen Relationen zwischen *ryōkoku/Ryōgoku* und Schlüsselwort. Denn selbst wenn *Ryōgoku*, *Ōzumō Hatsubasho* und *Kokugikan* in einem Satz auftauchen, könnte es sein, daß ersteres als Appellativum interpretiert werden muß. Das bloße Auffinden von Schlüsselwörtern kann zu Fehlurteilen führen (YOSHIMI 1997: 10).

Würde die angesprochene Person dieser Instruktion folgen, würde sie sich außen auf der Kühlerhaube sitzend wiederfinden. In englisch- und deutschsprachigen Ländern setzt man sich nämlich hinter das Lenkrad.

Pragmatische Aspekte, die im Japanischen vor allem den Ausdruck von Respekt und Höflichkeit, den sozialen Status und das Geschlecht des Rezipienten betreffen, sind bei der maschinellen Übersetzung ins Japanische von größerer Bedeutung als bei der maschinellen Übersetzung vom Japanischen in europäische Sprachen. Da ohnehin davon ausgegangen wird, daß die heute gängigen maschinellen Übersetzungssysteme zum einen pragmatische Aspekte nicht berücksichtigen können und zum anderen nur für die Übersetzung terminologisch eng determinierter Bereiche geeignet sind, ist die hiermit in Zusammenhang stehende Problematik auf Diskussionen rein theoretischer Natur beschränkt.²⁰

Die vorangegangenen Ausführungen haben deutlich gemacht, daß die Qualität von maschinellen Übersetzungssystemen, die aus dem Japanischen in europäische Sprachen übersetzen, starken lexikalischen, syntaktischen und semantischen Beschränkungen unterworfen ist. Da solche Systeme weder Kontextwissen verarbeiten noch von anderen Wissensquellen als denen linguistischer Natur profitieren können, ist davon auszugehen, daß eine qualitativ hochwertige maschinelle Übersetzung ohne menschliches Eingreifen derzeit nicht möglich ist und auf absehbare Zeit auch nicht möglich sein wird. Bei dem Output von maschinellen Übersetzungssystemen handelt es sich daher immer um „Rohübersetzungen“, die der Post- oder Präedition bzw. weiterer Modifikationen bedürfen (SPARCK JONES und GALLIERS 1996: 70–71). Der qualitativen Bewertung bereits existierender maschineller Übersetzungssysteme im kommerziellen Bereich – für potentielle Anwender und Käufer – und im Bereich der Forschung und Entwicklung – zur Verbesserung der Systeme – sind daher enge Grenzen gesetzt. Im Rahmen zahlreicher Projekte, die derzeit in den USA, Europa und Japan an der Entwicklung einer neuen Generation von maschinellen Übersetzungssystemen arbeiten, stehen neben den herkömmlichen, computerlinguistisch motivierten Evaluationsmethoden, die der qualitativen Verbesserung des Übersetzungsoutputs besonders auf lexikalischer und syntaktischer Ebene zugute kommen, neuerdings auch kognitive und übersetzungswissenschaftliche Überlegungen im Mittelpunkt der Forschungen. Hierbei wird untersucht, welche Strate-

²⁰ Es sei jedoch darauf hingewiesen, daß bei maschinellen Übersetzungssystemen vom Englischen ins Japanische in der Regel mindestens unter zwei Höflichkeitsebenen ausgewählt werden kann (höflichkeitsleere *ru/u*-Form oder standardhöfliche *desu/masu*-Form).

gien Humanübersetzer während des Übersetzungsprozesses anwenden und ob diese Strategien maschinell simulierbar sind.

Eine der ersten Methoden, die in Japan aus den Forschungen der neuen Generation hervorging, war die sogenannte *example-based machine translation* (EBMT) bzw. *machine translation by analogy*, die grundlegend davon ausgeht, daß bestimmte Satzmuster im Gedächtnis gespeichert sind, die mit identischer syntaktischer Struktur, aber unterschiedlichen Wörtern immer wiederkehren und ohne Kenntnis grammatischer Regeln immer wieder auf die gleiche Weise übersetzt werden können (FUJII 1996: 17–18, SHIRAI 1997: 8, WATANABE 1998: 10–13). Ein maschinelles Übersetzungssystem, das Textdaten ohne lexikalische und syntaktische Analyse gewinnen kann, wird *translation memory* genannt. Bei der Überarbeitung von Gebrauchsanweisungen, Handbüchern u.ä. ändert sich oft nur ein Bruchteil der Produktbeschreibungen gegenüber der ursprünglichen Version. Von Verbesserungen eines Produktes sind durchschnittlich nur zehn bis zwanzig Prozent des Gesamttextes der jeweiligen Gebrauchsanleitung betroffen (SHIRAI 1997: 8). Bei der Überarbeitung der für den Export bestimmten Übersetzungen der Gebrauchsanweisungen und Handbücher greifen maschinelle Übersetzungssysteme der neuen Generation per *translation memory* auf die bereits vorliegenden Übersetzungen zurück. Da maschinelle Übersetzungssysteme überdurchschnittlich häufig im sogenannten *manual*-Bereich zum Einsatz kommen, wird durch diese neue Methode viel Zeit und Geld gespart.

Wie bereits angesprochen, ist es auf absehbare Zeit unmöglich, einem Computer Weltwissen bzw. Common sense zu vermitteln, da der damit verbundene Arbeitsaufwand zur Erstellung semantischer Netze zu hoch ist. Die *example-based translation* bietet der maschinellen Übersetzung aber immerhin die Möglichkeit, Sätze korrekt zu übersetzen, die ausschließlich mit Hilfe der lexikalischen, syntaktischen und semantischen Analyse nicht korrekt übersetzt werden könnten. Damit diese Methode in der Praxis eingesetzt werden kann, werden seit einigen Jahren große Beispielsatz-Korpora angelegt.²¹ Je umfangreicher solche Textkorpora sind, desto größer ist die Chance, zukünftig Erfolge mit statistischen Methoden für die maschinelle Übersetzung erzielen zu können (SHIRAI 1997: 11).²² Bei-

²¹ Ein umfangreicher japanischer Textkorpus ist z.B. der *JEIDA Nichi-Ei taiyaku kōpasu* unter Leitung der Japan Electronics Industries Development Association (JEIDA), der Übersetzungsäquivalente aus japanischen Weißbüchern speichert und inzwischen einen Umfang von mehreren zehntausend Sätzen hat (ISAHARA 1998: 2–5).

²² Zur Kritik an statistischen und beispielsatzorientierten Methoden der maschinellen Übersetzung vgl. TANAKA (1996: 2–3).

spielsatz-Korpora dienen aber nicht nur als Basis für verschiedene korpusbasierte Methoden, sondern auch als Testkorpora für linguistische Evaluationen von maschinellen Übersetzungssystemen.

Eine japanische Erhebung von 1994/95 hat ergeben, daß der Bedarf an maschineller Übersetzungssoftware in Japan fast ebenso hoch ist wie die Nachfrage nach Textverarbeitung (ATSUKAWA 1996: 12). Es stellt sich die Frage, warum dieser extrem hohen Nachfrage bis heute nicht zufriedenstellend entsprochen werden kann. Neben den Faktoren Preis, Übersetzungsgeschwindigkeit, Entwicklungszeit für die Generierung von englischem Text und angemessene Berücksichtigung der Anwenderbedürfnisse – Faktoren, die seit der Kommerzialisierung der ersten japanischen maschinellen Übersetzungssysteme sprunghaft verbessert werden konnten und zusammen mit der raschen Verbreitung des PC das Anwenderspektrum erweitert und die Marktsituation verändert haben – spielt die Übersetzungsqualität dabei die entscheidende Rolle. Bei Entwicklern und Herstellern steht die Verbesserung der Übersetzungsqualität heute zwar an erster Stelle, und man ist sich darüber im klaren, ein in der Entwicklung unfertiges, fehlerhaftes Produkt zu vertreiben. Aber aufgrund der den gängigen maschinellen Übersetzungssystemen zugrunde liegenden veralteten Technologie werden kleine Verbesserungen heute nur durch Erweiterung und Veränderung der Lexika und Grammatiken erzielt. Eine wirklich bemerkenswerte Steigerung der Übersetzungsqualität kann – wenn überhaupt – nur vor dem Hintergrund wirklich neuer Entwicklungslinien in der Forschung und Entwicklung und somit nicht kurz- und mittelfristig, sondern nach jahrzehntelangem Forschungsaufwand erwartet werden. Mittelfristige Verbesserungen in der Übersetzungsqualität können mit einer Kombination syntaktischer und beispielsatzorientierter Methoden erwartet werden, wobei einfach zu übersetzende Sätze mit den herkömmlichen Strategien, schwer übersetzbare Sätze mit Hilfe von EBMT übersetzt werden.

Der heutige japanische Fokus in Forschung und Entwicklung auf den sogenannten korpusbasierten Methoden der maschinellen Übersetzung (EBMT und statistische maschinelle Übersetzung) sollte jedoch nicht von der Notwendigkeit ablenken, sich mit den tieferen Strukturen der menschlichen Sprache auseinanderzusetzen. Dies erfordert, neben einer in Zukunft – besonders in Japan – wesentlich intensiveren multidisziplinären Zusammenarbeit mit Linguistik, Übersetzungswissenschaft und Kognitionsforschung, neue Erkenntnisse über die Verarbeitung menschlicher Sprache im Gehirn, die Entwicklung neuer, semantikorientierter Methoden sowie einen starken Forschungsschwerpunkt auf der linguistischen Evaluation von maschinellen Übersetzungssystemen.

LITERATURVERZEICHNIS

- AAMT (1996): Doko made dekiru? Konpyūta hon'yaku [Bis wohin ist die Computerübersetzung möglich?]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 17 (December 1996). Tōkyō: AAMT, S. 21.
- AAMT (1998): Kikai hon'yaku kanren sofutouea ichiranhyō [Aufstellung über maschinelle Übersetzungssoftware]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 23 (July 1998). Tōkyō: AAMT, S. 18–23.
- ALPAC (Hg.) (1966): *Languages and Machines: Computers in Translation and Linguistics*. Report by the Automatic Language Processing Advisory Committee, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council (Publication 1416).
- ATSUKAWA Yoshikazu (1996): Hon'yaku sofuto shijō no genzai [Der gegenwärtige Markt für Übersetzungssoftware]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 16 (September 1996). Tōkyō: AAMT, S. 12–13.
- CARBONELL, Jaime (1992): Machine Translation in Japan and the U.S. In: JAPANESE TECHNOLOGY EVALUATION CENTER (JTEC) (Hg.): *JTEC Panel Report on Machine Translation in Japan*. Baltimore, Maryland: Loyola College, S. 7–22.
- FUJI Masaru (1996): Kikai hon'yaku to bunpō [Maschinelle Übersetzung und Grammatik]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 16 (September 1996). Tōkyō: AAMT, S. 16–20.
- FUJINAMI, Tsutomu und Christine NANZ (1997): *The 101 Translation Problems between Japanese and German/English* (Vm-Report 199, July 1997). Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- HUTCHINS, W. John (1986): *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood.
- HUTCHINS, W. John und Harold L. SOMERS (1992): *An Introduction to Machine Translation*. London, San Diego: Academic Press.
- ISAHARA Hitoshi (1998): Tagengo kōpasu no dōkō: Denshi kyō Nichi-Ei taiyaku kōpasu [Trends bei multilingualen Korpora: Elektronischer Korpus für Japanisch-Englisch]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 23 (July 1998). Tōkyō: AAMT, S. 2–5.
- KAMIGANE Kōhei (1997): CICC kinrin shokoku aida no kikai hon'yaku shisutemu purojekuto [Das CICC-Projekt zur Entwicklung eines maschinellen Übersetzungssystems für die Übersetzung zwischen Spra-

- chen von Nachbarstaaten]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 19 (June 1997). Tōkyō: AAMT, S. 29–33.
- KAWASHIMA Makoto (1997): CICC tagengo kikai hon'yaku shisutemu forō appu jigyō [Folgeprojekt für das multilinguale CICC-MÜ-System]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 18 (March 1997). Tōkyō: AAMT, S. 18–21.
- KOGA Katsuo (1996): Hon'yaku sofuto no shijō dōkō [Tendenzen des Marktes für Übersetzungssoftware]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 15 (June 1996). Tōkyō: AAMT, S. 6–8.
- KOIZUMI Atsuko (1996): Kikai hon'yaku to jisho [Maschinelle Übersetzung und Lexika]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 17 (December 1996). Tōkyō: AAMT, S. 10–15.
- KOLLER, Werner (1997): *Einführung in die Übersetzungswissenschaft*. 5., aktualisierte Auflage. Wiesbaden: Quelle und Meyer.
- KUMANO Akira (1996): Kikai hon'yaku to hinshi [Maschinelle Übersetzung und Wortklassen]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 15 (June 1996). Tōkyō: AAMT, S. 18–23.
- LEHRBERGER, John und Laurent BOURBEAU (1988): *Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam, Philadelphia: Benjamins.
- LEÓN, Marjorie und Julia AYMERICH (1997): *Making MT Work for You*. MT Summit VI, 29 October – 1 November 1997, San Diego, California (unveröffentlichter Vortrag).
- METZING, Dieter und Melanie SIEGEL (1994): *Zero pronoun processing: some requirements for a Verbmobil system* (Vm-Memo 46, September 1994). Bielefeld: Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft, Abt. Computerlinguistik.
- NAGAO Makoto (1989): *Machine Translation: how far can it go?* Übersetzt von Norman D. Cook. Oxford: Oxford University Press.
- NAGAO Makoto und MAKINO Takenori (Hg.) (1995): *Konpyūta de hon'yaku suru* [Mit Computern übersetzen]. Tōkyō: Kyōritsu shuppan.
- NARITA Hajime (Hg.) (1994): *Kō sureba tsukaeru kikai hon'yaku* [Wie man die maschinelle Übersetzung anwendbar machen kann]. Tōkyō: Baberu Puresu.
- NARITA Hajime (1997): *Pasokon hon'yaku no sekai* [Die Welt der PC-Übersetzung]. Tōkyō: Kōdansha.
- NIRENBURG, Sergei, Jaime CARBONELL, Masaru TOMITA und Kenneth GOODMAN (1992): *Machine Translation: A Knowledge-Based Approach*. San Mateo, CA: Morgan Kaufmann.

- SCHWANKE, Martina (1991): *Maschinelle Übersetzung: Ein Überblick über Theorie und Praxis*. Berlin und Heidelberg: Springer.
- SHIRAI Satoshi (1997): Taiyaku reibun to kikai hon'yaku [Zweisprachige Beispielsätze und maschinelle Übersetzung]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 20 (September 1997). Tōkyō: AAMT, S. 7–12.
- SHU Gyokuryu (1997): Nihongo no benkyō to hon'yaku [Japanischstudium und -übersetzung]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 19 (June 1997). Tōkyō: AAMT, S. 21–24.
- SIEGEL, Melanie (1994): *Definitheit und Numerus: Anforderungen an den Transfer Japanisch-Englisch* (Vm-Memo 56, Dezember 1994). Bielefeld: Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft, Abt. Computerlinguistik.
- SIEGEL, Melanie (1997): *Die maschinelle Übersetzung aufgabenorientierter japanisch-deutscher Dialoge: Lösungen für Translation Mismatches*. Berlin: Logos.
- SPARCK JONES, Karen und Julia R. GALLIERS (1996): *Evaluating Natural Language Processing Systems: An Analysis and Review* (Lecture notes in computer science, 1083). Berlin, Heidelberg u.a.: Springer. (Vorher erschienen als: Technical Report 291, University of Cambridge Computer Laboratory, March 1993).
- TANAKA Hozumi (1996): Kikai hon'yaku shisutemu no kongo [Die Zukunft von MÜ-Systemen]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 16 (September 1996). Tōkyō: AAMT, S. 2–4.
- VASCONCELLOS, Muriel (1992): Languages and Application Domains. In: JAPANESE TECHNOLOGY EVALUATION CENTER (JTEC) (Hg.): *JTEC Panel Report on Machine Translation in Japan*. Baltimore, Maryland: Loyola College, S. 41–46.
- WATANABE Hiyoshi (1998): Yōrei o mochiita hon'yaku [Beispielsatzorientierte Übersetzung]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 22 (April 1998). Tōkyō: AAMT, S. 10–13.
- YOSHIMI Katsuhiko (1997): Ningen no kizukanai aimaisa [Vom Menschen unbemerkte Ambiguitäten]. In: AJIA TAIHEIYŌ KIKAI HON'YAKU KYŌKAI (AAMT) (Hg.): *AAMT Journal* 18 (March 1997). Tōkyō: AAMT, S. 7–10.